



Science & Technology  
Facilities Council

# CCMVal Archive at BADC

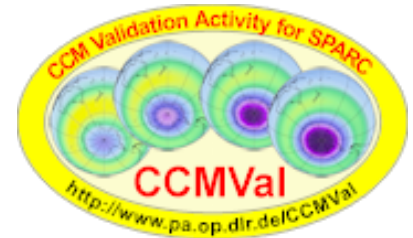
Charlotte Pascoe

Bryan Lawrence, Alison Pamment, David Hooper, Sam Pepler



**British Atmospheric  
Data Centre**

NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE  
NATURAL ENVIRONMENT RESEARCH COUNCIL



# CCMVal Archive at BADC

- BADC
- CF-NetCDF
- Three good reasons to keep Data:
  - Reuse: in 10 minutes, in 10 years...
  - Repurposing: discovery
  - Citation: data journals - OJIMS
- CMIP5 and Metafor: Metadata Capture and YOU!

# CEDA

## Centre for Environmental Data Archival

NERC designated data centres for Atmospheric Science and EO



IPCC  
official data centre  
<http://ipcc-data.org>

- British Atmospheric Data Centre
- NERC Earth Observation Data Centre
- IPCC Data Distribution Centre
- UK Projects
  - NDG (NERC Data Grid)
  - C-SEKT (New CSML and CSML-WFS)
  - OJMS (Data Publication)
  - UKCIP (UK-Scale Climate Predictions)
- EU Projects
  - Metafor (Climate Model Documentation)
  - EUFAR (Distributed Archive for European Aircraft)
  - IS-ENES (Distributed Archive for European Climate Model Data)

# ... and just for the BADC

- BADC has approximately 150 real datasets (and thousands of virtual datasets).
- BADC has tens of millions of files containing thousands of measured or simulated parameters.

• BADC the with describe ong em.

• Calend

**CCMVal Data:  
3.4 TB**

- 1
- downloaded 30 TB data
- in 13 million files
- from 134 datasets.

***Less than half of the BADC data consumers are “atmospheric science” users!***





Facilitating the exchange of data between modelling groups

## Standard Names

Each physical, chemical or biological quantity is given a CF standard name and assigned appropriate units

The purpose of the standard name is to determine whether the parameters from different data sources can be treated as directly comparable.

## Coordinate Variables

CF metadata describes multi-dimensional data by attaching a coordinate variable to each dimension

## Time

A separate variable is used to describe the time dimension

Standard Name	Canonical Units	AMIP	GRIB
air_potential_temperature	K	theta	13
air_pressure	Pa	plev	1

# CF-NetCDF

Files composed of just three basic entry types

## Global Attributes

store items of metadata which are relevant to the file as a whole

## Dimensions

e.g. **time**, **altitude**, **range**

each one has a corresponding coordinate **variable** which stores the coordinate values along the dimension

## Variables

contain the actual data

can have **Attributes** attached to them

The first 10 (of 240) lines from the file **sw010203**  
(taken from the NERC MST Radar Facility archives)

4.31	155.3	3.92	136.1	5.15	140.2	4.23	137.1	4.75	150.2	4.71	137.9
4.35	146.5	4.52	138.0	4.83	153.7	5.40	145.8	4.63	141.0	4.90	137.3
4.31	143.3	4.58	157.0	4.94	141.7	4.65	143.1	4.63	143.0	4.88	149.5
5.42	148.5	4.92	140.4	4.04	146.7	3.92	151.5	5.02	135.3	5.06	151.6
4.65	152.3	4.31	168.8	3.79	145.3	5.92	152.9	5.02	145.8	4.77	161.6
4.79	144.1	4.60	147.5	5.33	150.1	4.81	141.0	6.02	146.9	4.38	149.0
4.42	142.5	4.58	133.4	4.35	150.5	4.96	149.8	5.56	143.4	5.08	148.5
5.19	141.6	4.40	142.4	4.10	152.6	5.02	134.0	4.94	142.9	5.27	144.4
5.38	141.5	5.88	144.8	6.00	140.1	4.75	158.3	5.08	148.1	5.46	163.5
4.27	150.8	4.69	138.8	5.71	144.0	5.21	138.8	5.00	132.4	5.06	144.4

## What is known about this file?

**sw** indicates that the file contains "surface" wind data  
(i.e. speed and direction) from the location Frongoch

**010203** represents the date in YYMMDD format

1st February 2003  
(British convention)

2nd January 2003  
(North American convention)

3rd February 2001  
(Swedish convention)

The first 10 (of 240) lines from the file **sw010203**  
(taken from the NERC MST Radar Facility archives)

4.31	155.3	3.92	136.1	5.15	140.2	4.23	137.1	4.75	150.2	4.71	137.9
4.35	146.5	4.52	138.0	4.83	153.7	5.40	145.8	4.63	141.0	4.90	137.3
4.31	143.3	4.58	157.0	4.94	141.7	4.65	143.1	4.63	143.0	4.88	149.5
5.42	148.5	4.92	140.4	4.04	146.7	3.92	151.5	5.02	135.3	5.06	151.6
4.65	152.3	4.31	168.8	3.79	145.3	5.92	152.9	5.02	145.8	4.77	161.6
4.79	144.1	4.60	147.5	5.33	150.1	4.81	141.0	6.02	146.9	4.38	149.0
4.42	142.5	4.58	133.4	4.35	150.5	4.96	149.8	5.56	143.4	5.08	148.5
5.19	141.6	4.40	142.4	4.10	152.6	5.02	134.0	4.94	142.9	5.27	144.4
5.38	141.5	5.88	144.8	6.00	140.1	4.75	158.3	5.08	148.1	5.46	163.5
4.27	150.8	4.69	138.8	5.71	144.0	5.21	138.8	5.00	132.4	5.06	144.4

## What can we guess?

- Values are clearly arranged in pairs
  - 1st value of pair (e.g. 4.31) must represent speed - probably in units of  $\text{m s}^{-1}$
  - 2nd value of pair (e.g. 155.3) must represent direction - probably in units of  $^{\circ}$  from North (but meteorological or vector convention?)
- 240 lines, each with 6 columns, each with a pair of values  $\Rightarrow$  1440 pairs of values
- There are 1440 minutes in a day  $\Rightarrow$  1 minute sampling



The first 10 (of 240) lines from the file **sw010203**  
(taken from the NERC MST Radar Facility archives)

4.31	155.3	3.92	136.1	5.15	140.2	4.23	137.1	4.75	150.2	4.71	137.9
4.35	146.5	4.52	138.0	4.83	153.7	5.40	145.8	4.63	141.0	4.90	137.3
4.31	143.3	4.58	157.0	4.94	141.7	4.65	143.1	4.63	143.0	4.88	149.5
5.42	148.5	4.92	140.4	4.04	146.7	3.92	151.5	5.02	135.3	5.06	151.6
4.65	152.3	4.31	168.8	3.79	145.3	5.92	152.9	5.02	145.8	4.77	161.6
4.79	144.1	4.60	147.5	5.33	150.1	4.81	141.0	6.02	146.9	4.38	149.0
4.42	142.5	4.58	133.4	4.35	150.5	4.96	149.8	5.56	143.4	5.08	148.5
5.19	141.6	4.40	142.4	4.10	152.6	5.02	134.0	4.94	142.9	5.27	144.4
5.38	141.5	5.88	144.8	6.00	140.1	4.75	158.3	5.08	148.1	5.46	163.5
4.27	150.8	4.69	138.8	5.71	144.0	5.21	138.8	5.00	132.4	5.06	144.4

**In which order should we read the data?**

Column by column and then row by row or *vice versa*?

Try both ways and plot time series of the speed and direction data

There should be no sharp discontinuities in speed or direction

Vector (i.e. towards which the wind is blowing) or meteorological direction?

Compare with synoptic pressure maps or MST radar data

The first 10 (of 240) lines from the file **sw010203**  
(taken from the NERC MST Radar Facility archives)

4.31	155.3	3.92	136.1	5.15	140.2	4.23	137.1	4.75	150.2	4.71	137.9
4.35	146.5	4.52	138.0	4.83	153.7	5.40	145.8	4.63	141.0	4.90	137.3
4.31	143.3	4.58	157.0	4.94	141.7	4.65	143.1	4.63	143.0	4.88	149.5
5.42	148.5	4.92	140.4	4.04	146.7	3.92	151.5	5.02	135.3	5.06	151.6
4.65	152.3	4.31	168.8	3.79	145.3	5.92	152.9	5.02	145.8	4.77	161.6
4.79	144.1	4.60	147.5	5.33	150.1	4.81	141.0	6.02	146.9	4.38	149.0
4.42	142.5	4.58	133.4	4.35	150.5	4.96	149.8	5.56	143.4	5.08	148.5
5.19	141.6	4.40	142.4	4.10	152.6	5.02	134.0	4.94	142.9	5.27	144.4
5.38	141.5	5.88	144.8	6.00	140.1	4.75	158.3	5.08	148.1	5.46	163.5
4.27	150.8	4.69	138.8	5.71	144.0	5.21	138.8	5.00	132.4	5.06	144.4

It is often possible to "decode" ASCII files in this way, it is much more difficult for binary.  
No-one will be prepared to make this effort unless they have a strong need for the data.  
The data will become useless if the file name is changed - the date information is not recorded anywhere else.

Even if the data can be read, they may be of little scientific value unless something is known about: the type of instrument used, where it was located & how it was operated.

# The partial contents of file nerc-mstrf-wind-sensors\_capel-dewi\_20080114\_wxt510.nc

global attributes:

```
:verbose_metadata = "Free text description" ;  
:file_version_number = 1s ;  
:data_year = 2008s ;  
:data_month = 1s ;  
:data_day = 14s ;
```

dimensions:

```
time = 1440 ;
```

variables:

```
float longitude() ;  
    longitude:units = "degrees_east" ;  
    longitude:axis = "X"  
float latitude() ;  
    latitude:units = "degrees_north" ;  
    latitude:axis = "Y" ;  
float altitude() ;  
    altitude:units = "m" ;  
    altitude:axis = "Z" ;  
int time(time) ;  
    time:units = "seconds since 2008-01-14 00:00:00 +00:00" ;  
    time:axis = "T" ;  
float mean_wind_speed(time) ;  
    mean_wind_speed:units = "m s-1" ;  
    mean_wind_speed:coordinates = "latitude longitude altitude" ;  
    mean_wind_speed:cell_methods = "time: minimum (interval: 3 s)" ;  
    mean_wind_speed:missing_value = 99.9f ;  
short mean_wind_direction(time) ;  
    mean_wind_direction:units = "degree" ;  
    mean_wind_direction:coordinates = "latitude longitude altitude" ;  
    mean_wind_direction:cell_methods = "time: minimum (interval: 3 s)" ;  
    mean_wind_direction:missing_value = 999s ;
```

# We need metadata

We will never *look* at all our data. We need to process it on ingestion.

We *have* to provide tools so users can select what they want

**Tools need metadata!**

We can't manually reprocess our files to create new information about the data we hold, we *have* to automate

**Automation needs compliant metadata**


**Discovery Choosing Manipulating**

All need input metadata to drive tools

**Discovering  
Rain**



# Ontology Mediated Discovery



## Data Discovery Service

Search Results

New text search  All

92 results for "precipitation" (Global)

You could also try - narrower search(s): [hail](#) [rain](#) [snow](#) ; broader search(s): [meteorology](#) .

[Refine search](#)  | Found 92 | Showing 1-10 | [Next 10](#)

Dataset description	Temporal coverage		Spatial coverage
	Start date	End date	
<b>Citation:</b> Esch, Monika (2005): ECHAM4_OPYC_SRES_B2: 110 YEARS COUPLED B2 RUN 6H VALUES <b>Title:</b> ECHAM4_OPYC_SRES_B2: 110 YEARS COUPLED B2 RUN 6H VALUES <b>Abstract:</b> The SRES data sets were published by the IPCC in 2000 and classified into four different scenario families (A1, A2, B1, B2). SRES_B2 storyline describes a world in which the emphasis is on local solut <a href="#">more</a> <b>Sourced from:</b> <a href="#">WDCC</a> <b>Links:</b> <a href="#">  </a> <a href="#">&gt;</a> <a href="#">🔗</a>	230-1-1	340-12-30	90.0 0.0 360.0 -90.0
<b>Citation:</b> Stendel, Martin; Schmith, Torben; Roeckner, Erich; Cubasch, Ulrich (2004): IPCC_ECHAM4OPYC_SRES_A2_MM <b>Title:</b> IPCC_ECHAM4OPYC_SRES_A2_MM <b>Abstract:</b> The SRES data sets were published by the IPCC in 2000 and classified into four different scenario families (A1, A2, B1, B2). SRES_A2 storyline describes a very heterogeneous world with the underlying <a href="#">more</a> <b>Sourced from:</b> <a href="#">WDCC</a> <b>Links:</b> <a href="#">  </a> <a href="#">&gt;</a> <a href="#">🔗</a>	1990-1-1	2100-12-31	90.0 0.0 360.0 -90.0
<b>Citation:</b> Esch, Monika (2005): IPCC_ECHAM4OPYC_SRES_B2_MM <b>Title:</b> IPCC_ECHAM4OPYC_SRES_B2_MM			90.0



# How discovery works

## Vocabulary server for CF standard names

Defines: **Exact** (for previous versions of the standard name), **Narrow** and **Broad** mappings



**Rain**.....

- Broad Match: Precipitation

**Precipitation**

- Narrow Match: Hail, Rain, Snow

- Broad Match: Meteorology

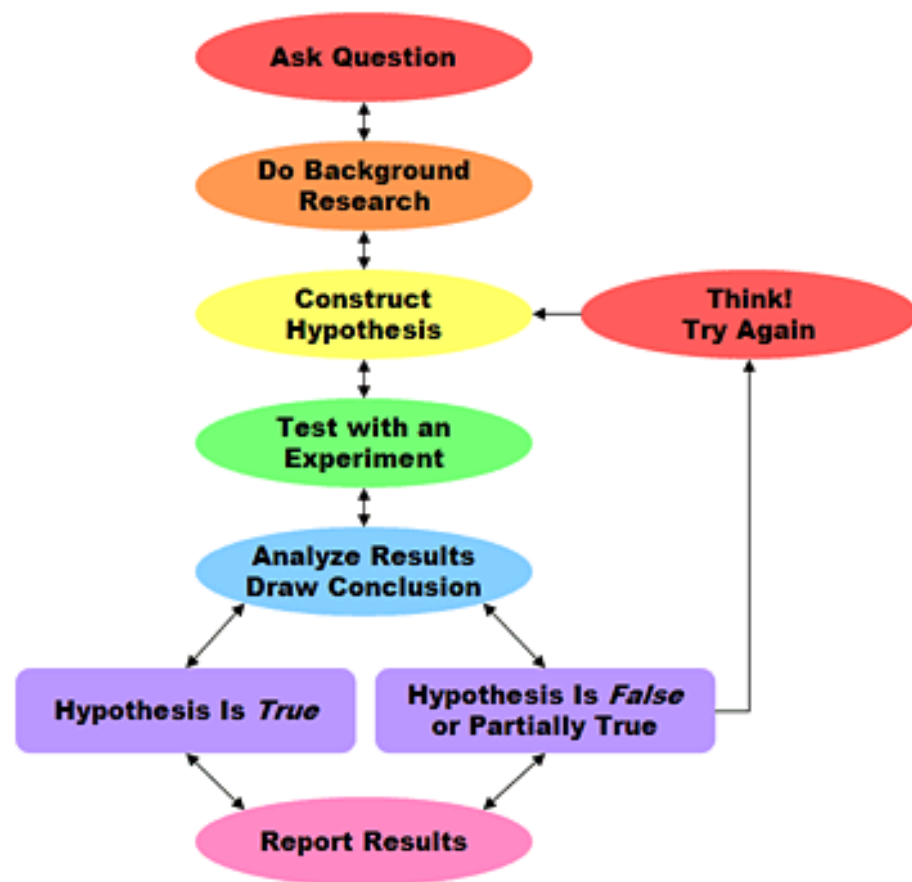
Building the mappings between vocabularies requires understanding the science.

Roy Lowry (BODC) has built the equivalences that are currently in the vocab server but...

**More people** will need to **get involved** if full mappings are to be provided between all the stored vocabularies.

# The Scientific Method

- depends on Analyze Results,
- finishes with REPORT RESULTS!
- Once upon a time, data were a few numbers, maybe a table, and the “paper” included “results”.
- That's very rarely true now, especially in environmental sciences!
- Do we care any more about the **evidential basis** for the scientific method? Do we care about **repeatability**?





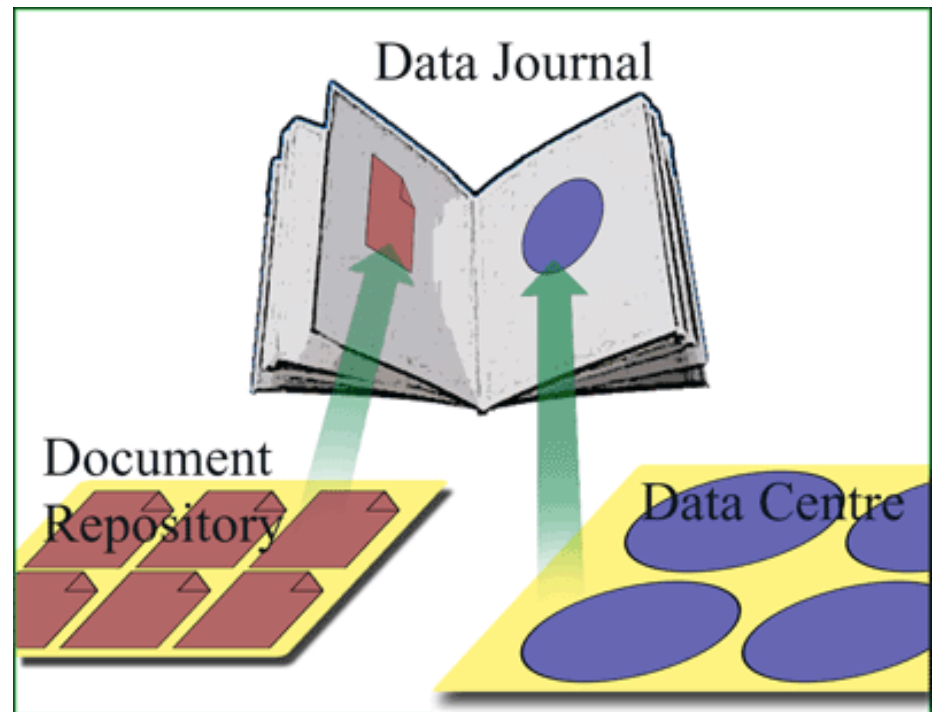
# Overlay Journal Infrastructure for Meteorological Sciences

<http://proj.badc.rl.ac.uk/ojims>

*Aid reusability:* make data  
**permanently** available and the  
knowledge within useable in other  
contexts

*Enable Recognition:* make it possible  
to measure and recognise the  
value of work

- Author prepares the overlay document.
- Author submits the dataset and the overlay document.
- Reviewer reviews the overlay document and the dataset




# CMIP5 Metadata Capture



Collect information from modelling groups:

- Details of models used
- Details of simulations carried out
- Details of how models conform to experiment requirements
- Details of hardware used
- Make information available to Curator, PCMDI, METAFOR and others



"finding and understanding simulations of past, present and future climate"

---

DRAFT CMIP5 Questionnaire

---

## NCAS Climate HIGEM Project

Each centre has associated with it a number of simulations, run with a model, and conforming to a CMIP5 experiment.

To add a simulation which uses a **new** model, you need to add the model first. Note that if the new model is a variant of an existing model, you can copy an old model and edit that!

To add an experiment which uses an old model, you can simply go straight to the add simulation option

### Experiments

This is the collection of CMIP5 experiments. For each experiment, you can see listed the simulations from this centre conforming to the model.

(NB: The experiments, and their descriptions will be entered via the admin interface, and there will be a lot of them !)

name	simulations
<a href="#">ESM PreInd Con</a>	<a href="#">control</a> <a href="#">Add New</a>
<a href="#">ESM hist</a>	<a href="#">Add New</a>
<a href="#">ESM RCP8.5</a>	<a href="#">Add New</a>
<a href="#">ESM fixed 1</a>	<a href="#">Add New</a>
<a href="#">ESM fixed 2</a>	<a href="#">Add New</a>
<a href="#">ESM feedback 1</a>	<a href="#">Add New</a>
<a href="#">ESM feedback 2</a>	<a href="#">Add New</a>
<a href="#">1% CO2</a>	<a href="#">Add New</a>
<a href="#">SST cont</a>	<a href="#">Add New</a>
<a href="#">CO2 Forcing</a>	<a href="#">Add New</a>
<a href="#">4xCO2 Abrupt</a>	<a href="#">Add New</a>

### Models and Platforms

You need to make sure you have at least one model and platform defined before you can create a simulation

[Add a new Model](#)
[Copy Model from another Centre](#)

#### Models associated with NCAS

HIGEM	<a href="#">edit</a>	<a href="#">CopyAsNew</a>
HIGEM2	<a href="#">edit</a>	<a href="#">CopyAsNew</a>

(CopyAsNew allows you to use an existing model as a template for a new one)

[Add a new Platform](#)

#### Computing platforms associated with NCAS

HECTOR	<a href="#">edit</a>
--------	----------------------

Icon Heaven

prototype of the questionnaire

For CMIP5, the questionnaire will be more flexible than the CCMVal prototype

Users can...

- complete the questionnaire in the order that suits them.
- add their own subcomponents.

- + GCM Template
- GCM Template
  - Atmosphere
    - AtmosAdvection
    - AtmosHorizontalDiffusion
    - AtmosVerticalDiffusion
    - + AtmosRadiation
    - + AtmosConvection
    - + AtmosClouds
    - AtmosPrecipitation
      - LargeScalePrecipitation
      - ConvectivePrecipitation
        - Microphysics
    - + AtmosGravityWaves

## Model Component LargeScalePrecipitation

Title

(Abbrev:  type:

Contact

### Sub Components

[Add Subcomponent](#)

### Parameters

Not all subcomponents will have parameters associated with them

Still got work to do to change these boxes to have drop down lists of vocab etc

Name	Value	Type
<input type="text" value="Scheme"/>	<input type="text"/>	<input type="text" value="XOR"/>
<input type="text" value="Type"/>	<input type="text"/>	<input type="text" value="XOR"/>
<input type="text" value="PrecipitatingHydrometeors"/>	<input type="text"/>	<input type="text" value="OR"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>

Not sure what to do about deletions. Do we want to allow them?

Should we have the parameters and subcomponents in the same box? How we handle them are quite different ... e.g. we update the form contents, or click on links ...

### References

[Modify references](#)

### Additional Information

Please tell us how this component has been improved from previous generations of this model:

Any additional information about this component:

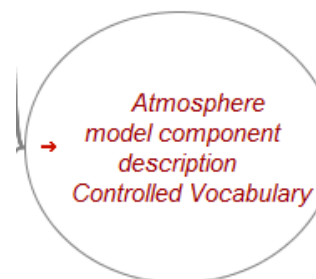
[update parameters and characteristics](#)

[Return to centre: NCAS](#)

Icon Heaven



# CMIP5 Metadata Capture



## Atmosphere

AtmosRadiation

AtmosConvection

AtmosClouds

AtmosPrecipitation

AtmosGravityWaves

### Large Scale Precipitation

Modelled

✗ Yes, Activiated

✗ Yes, Not Activiated

✗ No

SchemeName

✗ ⚠ <list of names>

✗ ⚠ Other

SchemeType

✗ ⚠ <list of names>

✗ ⚠ Other

PrecipitatingHydrometeors

✓ Liquid Rain

✓ Snow

✓ Hail

✓ Graupel

✗ ⚠ Other

### Convective Precipitation

Modelled

SchemeName

SchemeType

PrecipitatingHydrometeors

### Microphysics

Modelled

SchemeName

SchemeType

Properties



## Many controlled vocabularies being built: Software



Working closely with modelling community  
and Curator partners and..**YOU!**





Science & Technology  
Facilities Council

# Thank You



**British Atmospheric  
Data Centre**

NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE  
NATURAL ENVIRONMENT RESEARCH COUNCIL

Toronto, June 2009



# CCMVal Archive at BADC

- BADC
- CF-NetCDF
- Three good reasons to keep Data:
  - Reuse: in 10 minutes, in 10 years...
  - Repurposing: discovery
  - Citation: data journals - OJIMS
- CMIP5 and Metafor: Metadata Capture and YOU!